# Quantum ChromoDynamics (QCD)

J. Zinn-Justin*

*Dapnia, CEA/Saclay,*
*F-91191 Gif-sur-Yvette cedex, FRANCE,*

*and*

*Institut de Mathématiques de Jussieu–Chevaleret,*
*Université de Paris 7*

*email: zinn@dsm-mail.saclay.cea.fr

**CONTENT**



A reference:

J. ZINN-JUSTIN, *Quantum Field Theory and Critical Phenomena* (Oxford University Press, 1989) International Series of Monographs on Physics 113, 1054 pp. (2002), Fourth Edition

---

*Summation over repeated indices is always assumed in these lectures

## Geometry of gauge theories

*Gauge transformations and gauge invariance*

A gauge transformation is a local transformation: matter fields (here Dirac fermions) transform differently under a group at each point of space-time:

$$\psi'(x) = \mathbf{g}(x)\psi(x). \tag{1.1}$$

where the matrix $\mathbf{g}(x)$ belongs to a unitary representation $\mathcal{R}(G)$ of a Lie group $G$, *e.g.*, $SU(N)$.

For products of fields taken at the same point, global invariance ($\mathbf{g}$ constant) implies local invariance. This is no longer true for invariant functions of products of fields taken at different points. The problem can be solved by introducing parallel transporters $\mathbf{U}(C)$, which are oriented curve-dependent elements of $\mathcal{R}(G)$ that satisfy (below the extremity of $C_1$ coincides with the origin of $C_2$)

$$\mathbf{U}(C \equiv \{\cdot\}) = \mathbf{1}\,,\ \mathbf{U}(C^{-1}) = \mathbf{U}^{-1}(C)\ \ \mathbf{U}(C_1 \cup C_2) = \mathbf{U}(C_2)\mathbf{U}(C_1). \tag{1.2}$$

Then if $C$ is a curve joining point $y$ to $x$, one chooses the transformation of $\mathbf{U}(C)$ as

$$\mathbf{U}'(C) = \mathbf{g}(x)\mathbf{U}(C)\mathbf{g}^{-1}(y). \tag{1.3}$$

It then follows that the field

$$\psi_U = \mathbf{U}(C)\psi(y), \tag{1.4}$$

transforms by $\mathbf{g}(x)$ instead of $\mathbf{g}(y)$ and a quantity like $\bar{\psi}(x)\mathbf{U}(C)\psi(y)$ is gauge invariant.

Finally, for any closed curve $C$,

$$\mathrm{tr}\, U(C)$$

is gauge invariant.

This construction is valid both for continuum and discretized space.

In the continuum, in the limit of an infinitesimal differentiable curve, $y_\mu = x_\mu + \mathrm{d}x_\mu$, one can parametrize $\mathbf{U}(C)$ in terms of the connection or gauge field $\mathbf{A}_\mu(x)$, a space-vector and a matrix (anti-Hermitian) belonging to the representation of the Lie algebra of $\mathcal{R}(G)$:

$$\mathbf{U}(C) = \mathbf{1} + \mathbf{A}_\mu(x)\mathrm{d}x_\mu + o\left(\|\mathrm{d}x_\mu\|\right) \;\Rightarrow\; \mathbf{U}(C) = \mathrm{P}\exp\left(\oint_C \mathbf{A}_\mu(x)\mathrm{d}x_\mu\right)$$

The transformation properties of $\mathbf{A}_\mu(x)$ are obtained by expanding equation (1.3) at first order in $\mathrm{d}x_\mu$,

$$\mathbf{A}'_\mu(x) = \mathbf{g}(x)\mathbf{A}_\mu(x)\mathbf{g}^{-1}(x) + \mathbf{g}(x)\partial_\mu\mathbf{g}^{-1}(x). \tag{1.5}$$

From the point of view of global transformations ($\mathbf{g}$ constant), the field $\mathbf{A}_\mu(x)$ transforms by the adjoint representation of the group $G$. However, $\mathbf{A}_\mu(x)$ is not a tensor for gauge transformations, the transformation being affine.

*Covariant derivative.* In the limit of an infinitesimal curve,

$$
\begin{aligned}
\psi_U &= \big(\mathbf{1} + \mathbf{A}_\mu(x)\mathrm{d}x_\mu\big)\big(\psi(x) + \partial_\mu\psi(x)\mathrm{d}x_\mu\big) + o\left(\|\mathrm{d}x_\mu\|\right) \\
&= \left(\mathbf{1} + \mathrm{d}x_\mu \mathbf{D}_\mu\right)\psi(x) + o\left(\|\mathrm{d}x_\mu\|\right) \ \text{ with } \mathbf{D}_\mu = \mathbf{1}\,\partial_\mu + \mathbf{A}_\mu \,.
\end{aligned}
$$

$\mathbf{D}_\mu$ is the covariant derivative, whose explicit form depends on the tensor on which it is acting.
The identity

$$
\mathbf{g}(x)\left(\mathbf{1}\,\partial_\mu + \mathbf{A}_\mu\right)\mathbf{g}^{-1}(x) = \mathbf{1}\,\partial_\mu + \mathbf{g}(x)\mathbf{A}_\mu(x)\mathbf{g}^{-1}(x) + \mathbf{g}(x)\partial_\mu\mathbf{g}^{-1}(x), \quad (1.6)
$$

shows that $\mathbf{D}_\mu$ is a tensor, since $\mathbf{D}'_\mu$, the transform of $\mathbf{D}_\mu$ under the gauge transformation (1.5), is (the products have to be understood as products of differential and multiplicative operators)

$$
\mathbf{D}'_\mu = \mathbf{g}(x)\mathbf{D}_\mu\,\mathbf{g}^{-1}(x). \quad (1.7)
$$

*Infinitesimal gauge transformations.* Setting,

$$\mathbf{g}(x) = \mathbf{1} + \omega(x) + o\left(\|\omega\|\right),$$

in which $\omega(x)$ belongs to the Lie algebra of $\mathcal{R}(G)$, one derives from equation (1.5) the form of the infinitesimal gauge transformation of the field $\mathbf{A}_\mu$,

$$-\delta \mathbf{A}_\mu(x) = \partial_\mu \omega + [\mathbf{A}_\mu, \omega] \equiv \mathbf{D}_\mu \omega\,. \tag{1.8}$$

The equation yields the form of the covariant derivative in the adjoint representation. One verifies:

$$\partial_\mu \omega' + \left[\mathbf{A}'_\mu, \omega'\right] = \mathbf{g}(x)\left\{\partial_\mu \omega + [\mathbf{A}_\mu, \omega]\right\}\mathbf{g}^{-1}(x),$$

in which $\mathbf{A}'_\mu$ is given by equation (1.5) and $\omega'$ by

$$\omega'(x) = \mathbf{g}(x)\omega(x)\mathbf{g}^{-1}(x).$$

*Curvature tensor.* The commutator of two covariant derivatives

$$\mathbf{F}_{\mu\nu}(x) = [\mathbf{D}_\mu, \mathbf{D}_\nu] = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + [\mathbf{A}_\mu, \mathbf{A}_\nu],$$

is no longer a differential operator. It is again an element of the Lie algebra of $\mathcal{R}(G)$ and transforms, as a consequence of equation (1.7), as

$$\mathbf{F}'_{\mu\nu}(x) = \mathbf{g}(x)\mathbf{F}_{\mu\nu}(x)\mathbf{g}^{-1}(x).$$

$\mathbf{F}_{\mu\nu}$ is a tensor, the curvature tensor, generalization of the electromagnetic field of QED. The curvature tensor is associated with parallel transport along an infinitesimal closed curve.

## Gauge invariant action

*Matter fields.* For fermions transforming by $\mathcal{R}(G)$, the action

$$\mathcal{S}_{\mathrm{F}}(\bar{\psi}, \psi) = -\int \mathrm{d}^d x \, \bar{\psi}(x) \, (\not{\mathbf{D}} + M) \, \psi(x),$$

is gauge invariant.

*Gauge field.* The simplest gauge invariant action $\mathcal{S}(\mathbf{A}_\mu)$ function of the gauge field $\mathbf{A}_\mu$ has the form

$$\mathcal{S}(\mathbf{A}_\mu) = -\frac{1}{4g^2} \int \mathrm{d}^d x \, \mathrm{tr} \, \mathbf{F}^2_{\mu\nu}(x). \tag{1.9}$$

The sign in front of the action takes into account that, with our definition, the matrix $\mathbf{F}_{\mu\nu}$ is anti-Hermitian.

This action can be obtained from the gauge invariant quantity $\mathrm{tr}\, U(C)$ in the limit of an an infinitesimal closed contour $C$ (Fig. 1):

$$\mathrm{tr} \, \mathrm{P} \exp\left( \oint_C \mathbf{A}_\mu(x) \mathrm{d}x_\mu \right) - \mathrm{tr} \, \mathbf{1} \sim \mathrm{tr} \, \left[ \mathbf{F}_{\mu\nu}(x) \epsilon^{(1)}_\mu \epsilon^{(2)}_\nu \right]^2 .$$

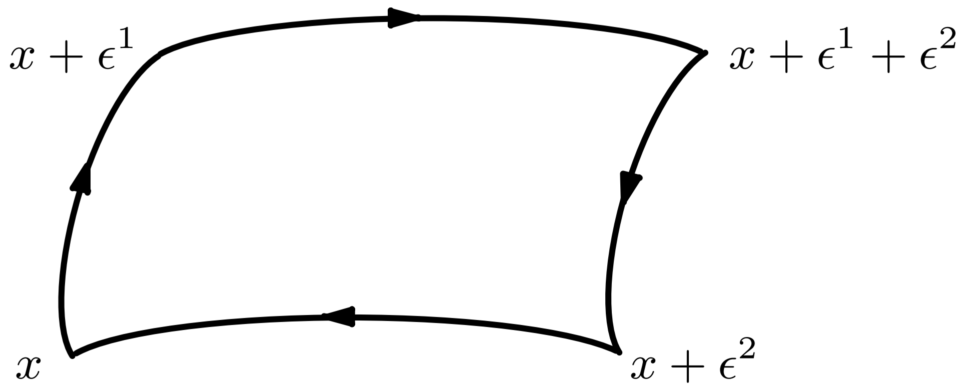


Fig. 1 The loop $C$.

Two remarks are immediately in order:

(i) In contrast with the Abelian case, because the gauge field transforms non-trivially under the group, (the gauge field is 'charged'), the curvature tensor $\mathbf{F}_{\mu\nu}$ is not gauge invariant, and thus not directly a physical observable. The action (1.9) is no longer a free field action; the gauge field has self-interactions and even the spectrum of the pure gauge action is non-perturbative (some analytic results can be obtained only in dimension two $(1+1)$).

Lattice gauge theory provides a framework for non-perturbative investigations.

(ii) As in the Abelian case, the action, because it is gauge invariant, does not provide a dynamics to the degrees of freedom of the gauge field which correspond to gauge transformations and, therefore, some gauge fixing is required. In the field integral language, the integral

$$\mathcal{Z} = \int [\mathrm{d}\mathbf{A}_\mu] \exp\left[\frac{1}{4g^2}\int \mathrm{d}^d x \operatorname{tr} \mathbf{F}^2_{\mu\nu}(x)\right]$$

is not defined because the integrand is constant along a gauge orbit.

## Hamiltonian formalism. Quantization in the temporal gauge

Non-Abelian gauge theories can be quantized in a simple way in the temporal or Weyl gauge, using a simple Hamiltonian formalism. This leads to a field theory that, at least at the formal level, is unitary because it corresponds to a Hermitian Hamiltonian. However, it lacks relativistic covariance and this is the source of many difficulties.

*Classical field equations*

In real time field theory, we denote by $t \equiv x_0 = ix_d$ time and the corresponding field component by $\mathbf{A}_0 = -i\mathbf{A}_d$. We use the notation $\dot{Q}$ for the time derivative of $Q$. Space components will carry roman indices $(\mathbf{A}_i, x_i)$.

To the real time form of the action (1.9) corresponds a classical field equation:

$$\mathbf{D}_\mu \mathbf{F}^{\mu\nu}(x) = 0\,., \tag{1.10}$$

The equation (1.10) does not lead to a standard quantization because the action does not depend on $\dot{\mathbf{A}}_0$, the time derivative of $\mathbf{A}_0$.

Thus, $\mathbf{A}_0$ is not a dynamical variable, the $\mathbf{A}_0$ field equation is a constraint equation that can be used to eliminate $\mathbf{A}_0$ from the action. However, in the absence of a mass term the reduced action does not depend on all space components of the gauge field. Only the combination $\left[\delta_{ij} - \mathrm{D}_i(\mathrm{D}_\perp^2)^{-1}\mathrm{D}_j\right]\dot{\mathbf{A}}_j$ appears ($\mathrm{D}_\perp^2$ is the covariant space Laplacian). But in contrast with the Abelian case the projector acting on $\mathbf{A}_i$ depends on the field itself, and, therefore, the procedure which led to Coulomb's gauge does not work here, at least in its simplest form. Therefore, it is simpler to begin with a quantization in the temporal (or Weyl) gauge.

*Temporal gauge.* Since any gauge transform of a solution to the field equations, is also a solution one can restrict fields by a gauge condition. One choice of gauge condition is specially well-suited to the construction of a Hamiltonian formalism (in particular, useful for finite temperature quantum field theory)

$$\mathbf{A}_0(t, x) = 0\,, \tag{1.11}$$

which defines the temporal gauge.

In this gauge, the field equations simplify and become (separating time and space components)

$$\dot{\mathbf{E}}_k = \mathbf{D}_l \mathbf{F}_{lk}, \quad \mathbf{D}_l \mathbf{E}_l = 0$$

with

$$\mathbf{E}_k = -\dot{\mathbf{A}}_k / g^2\,.$$

The first equation is a dynamical equation that can be directly derived from the initial Lagrangian in which the gauge condition has been used:

$$\mathcal{L}(\mathbf{A}_k) = -\mathrm{tr} \int \mathrm{d}^{d-1}x \left[ \frac{1}{2g^2} \dot{\mathbf{A}}_k^2(t,x) - \frac{1}{4g^2} \mathbf{F}_{kl}^2(t,x) \right]. \tag{1.12}$$

This expression defines a conventional Lagrangian for the space components of the gauge field: $\mathbf{E}_k$ is the conjugated momentum of $\mathbf{A}_k$.

The corresponding Hamiltonian is

$$\mathcal{H}(\mathbf{E},\mathbf{A}) = -\mathrm{tr} \int \mathrm{d}^{d-1}x \left[ \frac{g^2}{2} \mathbf{E}_k^2(x) + \frac{1}{4g^2} \mathbf{F}_{kl}^2(x) \right]. \tag{1.13}$$

By contrast, the second equation is a constraint equation, a non-Abelian generalization of Gauss's law. The only relevant solutions of the field equations are those that satisfy the constraint.

The constraint is compatible with the classical motion. Indeed, the gauge condition $\mathbf{A}_0 = 0$ is left invariant by time-independent gauge transformations:

$$-\delta \mathbf{A}_0(x) = \dot{\omega} + [\mathbf{A}_0, \omega] = 0 \text{ if } \dot{\omega} = 0\,.$$

Therefore, time-independent gauge transformations form a symmetry group of the Lagrangian (1.12) and thus of the Hamiltonian (1.13). The quantities $\mathbf{D}_l \mathbf{E}_l$ are the generators, in the sense of Poisson brackets, of the symmetry group.

These considerations immediately generalize to the quantum theory, the quantum operators $\mathbf{D}_l\mathbf{E}_l$ generators of a symmetry group, commute with the Hamiltonian. The space of admissible physical states $\Psi(\mathbf{A})$ is thus restricted by the quantum generalization of Gauss's law:

$$\mathbf{D}_l\mathbf{E}_l\Psi(\mathbf{A}) \equiv \mathbf{D}_l \frac{1}{i}\frac{\delta}{\delta\mathbf{A}_l(x)}\Psi(\mathbf{A}) = 0\,.$$

The equation implies that physical states are gauge invariant, that is, belong to the invariant sector of the symmetry group, a subspace which is left invariant by quantum evolution.

Quantization in the temporal gauge then follows conventional lines. Returning to the *Euclidean formalism*, one concludes that the partition function can be written as

$$\mathcal{Z} = \int [\mathrm{d}\mathbf{A}_\mu]\,\delta(\mathbf{A}_d)\exp\left[\frac{1}{4g^2}\int \mathrm{d}^d x\,\mathrm{tr}\,\mathbf{F}^2_{\mu\nu}(x)\right]. \qquad (1.14)$$

Note that at zero temperature the perturbative vacuum is automatically gauge invariant and Gauss's law plays no role, but this is no longer the case at finite temperature.

*Remarks.* The theory in the temporal gauge is not explicitly space-time covariant and this leads to serious difficulties. In particular, the theory is not renormalizable in the sense of power counting. Indeed the propagator in this gauge

$$W^{(2)}_{ij}(\mathbf{k}_\perp, k_d) = \frac{1}{k^2}\left(\delta_{ij} - \frac{k_i k_j}{\mathbf{k}_\perp^2}\right) + \frac{1}{k_d^2}\frac{k_i k_j}{\mathbf{k}_\perp^2},$$

in which $\mathbf{k}_\perp$ is the 'space' part of $\mathbf{k}$, does not decrease at $k_d$ fixed for large spatial momenta $|\mathbf{k}_\perp|$. (The poles in $\mathbf{k}$ lead also to difficulties.)

These problems are solved by showing that gauge invariant observables can equivalently be calculated from another quantum action which leads to a theory that is explicitly covariant and renormalizable by power counting.

*Covariant gauge*

By a set of transformations on the field integral that change correlation functions but not gauge invariant observables, one can pass to a covariant gauge constraining $\partial_\mu \mathbf{A}_\mu$. The partition function or vacuum functional $\mathcal{Z}$ then reads

$$\mathcal{Z} = \int \left[\mathrm{d}\mathbf{A}_\mu \, \mathrm{d}\bar{\mathbf{C}} \, \mathrm{d}\mathbf{C} \, \mathrm{d}\lambda\right] \exp\left[-\mathcal{S}(\mathbf{A}_\mu, \bar{\mathbf{C}}, \mathbf{C}, \lambda)\right], \tag{1.15}$$

where $\mathcal{S}$ is a local action:

$$\begin{aligned} \mathcal{S}(\mathbf{A}_\mu, \bar{\mathbf{C}}, \mathbf{C}, \lambda) = \int \mathrm{d}^d x \, \mathrm{tr}\Bigg[&-\frac{1}{4g^2}\mathbf{F}_{\mu\nu}^2 + \frac{\xi g^2}{2}\lambda^2(x) + \lambda(x)\partial_\mu \mathbf{A}_\mu(x) \\ &+ \mathbf{C}(x)\partial_\mu \mathbf{D}_\mu \bar{\mathbf{C}}(x)\Bigg], \end{aligned} \tag{1.16}$$

where $\bar{\mathbf{C}}$ and $\mathbf{C}$ spinless fermions, the Faddeev–Popov ‘ghosts’, and $\lambda$ a boson field, all transforming under the adjoint representation of the gauge group.

Except in the limit in which $\xi$ vanishes, it is also possible to integrate over $\lambda(x)$ to find a new quantum action:

$$\mathcal{S}(\mathbf{A}_\mu, \bar{\mathbf{C}}, \mathbf{C}) = \int \mathrm{d}^d x \,\mathrm{tr}\left\{-\frac{1}{g^2}\left[\frac{1}{4}\mathbf{F}^2_{\mu\nu} + \frac{1}{2\xi}\left(\partial_\mu \mathbf{A}_\mu\right)^2\right] + \mathbf{C}(x)\partial_\mu \mathbf{D}_\mu \bar{\mathbf{C}}(x)\right\}. \tag{1.17}$$

The obvious drawback of the covariant gauge, which leads to a covariant, local and renormalizable theory, is the lack of explicit positivity and thus unitarity. In particular, Faddeev–Popov fermions being spinless do not obey to the spin–statistics connection and are, therefore, unphysical.

*Remark.* As pointed out by Gribov, in contrast with the Abelian case, depending on the value of the gauge field $\mathbf{A}_\mu(x)$, the gauge condition

$$\partial_\mu \mathbf{A}_\mu(x) = \nu(x)$$

has not always a unique solution, a problem called Gribov's ambiguity.

When two solutions merge, the operator $\partial_\mu \mathbf{D}_\mu(\mathbf{A})$ has zero eigenvalues. The integral over the ghost fields $\mathbf{C}, \bar{\mathbf{C}}$ fields the factor

$$\int [\mathrm{d}\mathbf{C}\mathrm{d}\bar{C}] \exp\left[-\mathbf{C}(x)\partial_\mu \mathbf{D}_\mu \bar{\mathbf{C}}(x)\right] = \det \partial_\mu \mathbf{D}_\mu(\mathbf{A}).$$

(This is where the ghost action came about in the first place.)

This implies that the field integral representation of the gauge theory in the covariant gauge is not meaningful beyond perturbation theory. The same ambiguity has been shown to arise for a large class of gauge conditions.

*BRS symmetry*

One consequence of the covariant quantization procedure is that the quantized action is no longer gauge invariant. On the other hand the quantized action, in the covariant gauge, now has a BRS (Becchi–Rouet–Stora) symmetry, consequence of the stochastic dynamics given to the degrees of freedom of the gauge group variables. It is invariant under the fermion-like transformation (*cf.* supersymmetry)

$$\begin{cases} \delta \mathbf{A}_\mu(x) = -\varepsilon \mathbf{D}_\mu \bar{\mathbf{C}}(x), & \delta \bar{\mathbf{C}}(x) = \varepsilon \bar{\mathbf{C}}^2(x), \\ \delta \mathbf{C}(x) = \varepsilon \lambda(x), & \delta \lambda(x) = 0\,, \end{cases} \tag{1.18}$$

where $\varepsilon$ is a Grassmann generator.

One can also introduce the BRS differential operator

$$\mathcal{D} = \int \mathrm{d}^d x \,\mathrm{tr} \left[ -\mathbf{D}_\mu \bar{\mathbf{C}}(x) \frac{\delta}{\delta \mathbf{A}_\mu(x)} + \bar{\mathbf{C}}^2(x) \frac{\delta}{\delta \bar{\mathbf{C}}(x)} + \lambda(x) \frac{\delta}{\delta \mathbf{C}(x)} \right]. \tag{1.19}$$

This BRS operator is nilpotent (like a cohomology operator):

$$\mathcal{D}^2 = 0\,,$$

because two successive BRS transformations yields automatically zero.

The BRS symmetry of the quantized action is expressed by the equation

$$\mathcal{D}\mathcal{S}_{\text{gauge}}(\mathbf{A}_\mu, \bar{\mathbf{C}}, \mathbf{C}, \lambda) = 0\,,$$

($\mathcal{S}$ is BRS closed). Moreover, $\mathcal{S}_{\text{gauge}}$ is BRS exact, that is,

$$\mathcal{S}_{\text{gauge}} = \mathcal{D}\int \mathrm{d}^d x\, \mathrm{tr}\, \mathbf{C}(x)\left[\partial_\mu \mathbf{A}_\mu(x) + \xi g^2 \lambda(x)\right]. \tag{1.20}$$

Ward–Takahashi (WT) identities associated with the BRS symmetry take the form of a quadratic functional differential equation for the 1PI functional (the generating function of one-particle irreducible Feynman diagrams), symbolically

$$\Gamma * \Gamma = 0 \,.$$

This can be shown to imply the same equation for the renormalized action (Z.-J. 1974)

$$\mathcal{S}_{\mathrm{ren}} * \mathcal{S}_{\mathrm{ren}} = 0 \,.$$

The solution of this equation (a simple application of BRS cohomology methods) implies structural stability of the quantized action under renormalization.

The strategy generalizes to the renormalization of gauge invariant operators, but the method is less straightforward.

## Perturbation theory, regularization

Compared with the Abelian case, the new features of the non-Abelian case are the presence of gauge field self-interactions and ghost terms. In four dimensions, as in the Abelian case, the gauge field has dimension 1. The ghost fields has a simple $\delta_{ab}/p^2$ propagator and canonical dimension 1 in four dimensions. The interaction terms have all dimension 4 and, therefore, the theory is renormalizable by power counting in four dimensions. The power counting for matter fields is of course the same as in the Abelian case. Indeed, the coupling to matter fields differs from the Abelian case only by some geometric factors corresponding to group indices. For example, the coupling to fermions generated by the covariant derivative is simply $\gamma_\mu t^a_{ij}$.

*Infrared divergences.* In the covariant gauge, and in the absence of a Higgs mechanism that provides a mass to gauge fields, only the gauge $\xi = 1$, Feynman's gauge, leads to a theory which is obviously IR finite. In contrast to the Abelian case, it is impossible to give an explicit mass to the gauge field and to then construct a theory which is both unitary and renormalizable. On the other hand, one wants eventually to prove the gauge independence of the theory and therefore we must be able to define it for more than one gauge. One way to introduce an IR regulator is to consider the theory in a finite volume.

*Regularization.* Dimensional regularization is the most convenient for practical calculations and works in the absence of chiral fermions.

Lattice regularization, which is also relevant for non-perturbative calculations can be used generally since recently a method for handling chiral fermions has been discovered (related to Ginsparg–Wilson's relation).

Finally, momentum or Pauli–Villars's type regularizations work even in the chiral case but they regularize all diagrams except one-loop diagrams. The regularized pure gauge action takes the form:

$$\mathcal{S}(\mathbf{A}_\mu) = \int \mathrm{d}^d x \, \mathrm{tr}\, \mathbf{F}_{\mu\nu} P(\mathbf{D}^2/\Lambda^2) \mathbf{F}_{\mu\nu} \,,$$

the gauge function $\partial_\mu \mathbf{A}_\mu$ is changed into

$$\partial_\mu \mathbf{A}_\mu \longmapsto Q(\partial^2/\Lambda^2)\partial_\mu \mathbf{A}_\mu \,,$$

in which $P, Q$ are polynomials. In such a way both the gauge field propagator and the ghost propagator can be made arbitrarily convergent.

However, the covariant derivatives generate new interactions which are more singular. It is easy to verify that the power counting of one-loop diagrams is unchanged while higher order diagrams can be made convergent by taking the degrees of $P$ and $Q$ large enough.

For matter fields the situation is the same as in the Abelian case, for example, massive fermions contributions can be regularized by adding a set of regulator fields, massive fermions and bosons with spin.

Again in the case of chiral fermions, global chiral properties can be preserved, but problems arise with local chiral transformations. However, the problem of the compatibility between gauge symmetry and quantum corrections is reduced to an explicit verification of the WT identities for the one-loop diagrams. Note that the preservation of gauge symmetry is necessary for the cancellation of unphysical states in physical amplitudes and, thus, essential to the physical consistency of the quantum field theory.

*WT identities and renormalization.* From the BRS symmetry follow WT identities for correlation functions. They can be used to derive the form of the renormalized action. We give here the result only in the example of the pure gauge action in the covariant gauge, assuming that the gauge group $G$ is simple. Then the renormalized form of the action (1.17) is given by the substitution:

$$\begin{cases} g^2 \longmapsto Z_g g^2 \,, & \mathbf{A}_\mu \longmapsto Z_A^{1/2} \mathbf{A}_\mu \,, \\ \xi \longmapsto Z_A Z_g^{-1} \xi \,, & \mathbf{C}\bar{\mathbf{C}} \longmapsto Z_C \mathbf{C}\bar{\mathbf{C}} \,. \end{cases}$$

This result has a simple interpretation: the gauge structure is preserved and the coefficient of $(\partial_\mu \mathbf{A}_\mu)^2$ is unrenormalized exactly as in the Abelian case. However, unlike the Abelian case, the gauge transformation of the gauge field and, more generally the form of the covariant derivative, are modified by the gauge field renormalization.

## Quantum Chromodynamics: Renormalization group

Quantum Chromodynamics consists in a set of quarks characterized by a *flavour quantum number*, which are also triplets of a gauged symmetry, the $SU(3)$ *colour*, realized in the symmetric phase. Their interactions are mediated by the corresponding gauge fields (*gluons*):

$$\mathcal{S}(\mathbf{A}_\mu, \bar{\mathbf{Q}}, \mathbf{Q}) = -\int \mathrm{d}^4 x \left[\frac{1}{4g^2} \mathrm{tr}\, \mathbf{F}^2_{\mu\nu} + \sum_{\text{flavours}} \bar{\mathbf{Q}}_f \left(\not{\mathbf{D}} + m_f\right) \mathbf{Q}_f\right]. \tag{2.1}$$

The most important physical arguments in favour of such a model are

i) Quarks behave almost like free particles at short distances, as indicated by deep inelastic scattering experiments or the spectrum of bound states of heavy quarks. This is consistent with the sign of the RG $\beta$-function in non-Abelian gauge theories with not too many fermions. Moreover, according to the Coleman–Gross theorem, only non-Abelian gauge symmetries share this property called asymptotic freedom.

ii) No free quarks have ever been observed at large distance (but they manifest themselves indirectly in jet production). This is consistent with the simplest picture in which the $\beta$-function (which, due to AF, is negative at small coupling) remains negative for all couplings in such a way that the effective coupling constant grows without bounds at large distances. Numerical simulations strongly support this conjecture, called the confinement hypothesis.

## The Abelian anomaly

Anomalies arise when semi-classical symmetries of the theory cannot be implemented in the full quantum theory because some ordering of quantum operators is involved. An elementary example showing the role of operator ordering in implementing symmetries is provided by non-relativistic particle in a magnetic field

$$\hat{H} = \frac{1}{2m}\left[\hat{\mathbf{p}} + e\mathbf{A}(\hat{\mathbf{q}})\right]^2 .$$

Starting from the classical Hamiltonian, the problem of operator order arises in the product $\hat{\mathbf{p}} \cdot \mathbf{A}(\hat{\mathbf{q}})$. It is fixed by two independent conditions: gauge invariance and hermiticity. Fortunately, these two symmetries are consistent and yield the same result.

For the same reason, in gauge theories order in product of quantum operators is important. When gauge symmetry is confronted with fermion chiral symmetry, conflicts may appear that are called anomalies.

Technically, since none of the known regularization methods can deal in a straightforward way with one-loop diagrams in the case of chiral gauge fields (a feature related to operator ordering), this opens the possibility for anomalies.

We now show that indeed gauge theories with massless fermions and chiral symmetry can be found where the axial current is not conserved. The divergence of the axial current is then called an anomaly. This leads in particular to obstructions to the construction of gauge theories when the gauge field couples differently to the two fermion chiral components. Several examples are physically important like the constraint of anomaly cancellation in the theory of weak-electromagnetic interactions, the electromagnetic decay of the $\pi_0$ meson, or the $U(1)$ problem in QCD.

We first discuss the Abelian axial current and then the general non-Abelian case. The only possible source of anomalies are one-loop fermion diagrams in gauge theories when chiral properties are involved.

This reduces the problem to the discussion of fermions in the background of gauge fields, or equivalently to the properties of the determinant of the gauge covariant Dirac operator.

*Abelian axial current and Abelian vector gauge field*

We first consider the QED-like fermion action $\mathcal{S}(\bar{\psi},\psi)$ for massless Dirac fermions $\psi,\bar{\psi}$ in the background of an Abelian gauge field $A_\mu$:

$$\mathcal{S}(\bar{\psi},\psi)=-\int \mathrm{d}^4x\,\bar{\psi}(x)\not{D}\psi(x),\quad \not{D}\equiv\not{\partial}+ie\not{A}\,, \tag{2.2}$$

and the corresponding field integral

$$\mathcal{Z}(A_\mu)=\int\left[\mathrm{d}\psi\mathrm{d}\bar{\psi}\right]\exp\left[-\mathcal{S}(\psi,\bar{\psi})\right]=\det\not{D}\,. \tag{2.3}$$

In what follows we denote by $\langle\bullet\rangle$ expectation values with respect to the measure $\mathrm{e}^{-\mathcal{S}(\psi,\bar{\psi})}$.

One can find regularizations which preserve gauge invariance and, since the fermions are massless, chiral symmetry. One would, therefore, naively expect the corresponding axial current to be conserved. However, the proof of current conservation involves space-dependent chiral transformations and, therefore, steps that cannot be regularized without breaking one of the symmetries.

The coefficient of $\partial_\mu \theta(x)$ in the variation of the action under a space-dependent chiral transformation

$$\psi_\theta(x) = \mathrm{e}^{i\theta(x)\gamma_5}\psi(x), \qquad \bar{\psi}_\theta(x) = \bar{\psi}(x)\mathrm{e}^{i\theta(x)\gamma_5}, \tag{2.4}$$

yields the axial current $J^5_\mu(x)$. For the action (2.2) one finds,

$$\delta\mathcal{S} = \int \mathrm{d}^4 x\, \partial_\mu \theta(x) J^5_\mu(x) \quad \text{with} \quad J^5_\mu(x) = i\bar{\psi}(x)\gamma_5\gamma_\mu\psi(x). \tag{2.5}$$

After the transformation (2.4), $\mathcal{Z}(A_\mu)$ becomes

$$\mathcal{Z}(A_\mu, \theta) = \det\left[\mathrm{e}^{i\gamma_5\theta(x)} \not{D} \mathrm{e}^{i\gamma_5\theta(x)}\right]. \tag{2.6}$$

Since $\mathrm{e}^{i\gamma_5\theta}$ has a determinant which is unity, one would naively conclude that $\mathcal{Z}(A_\mu, \theta) = \mathcal{Z}(A_\mu)$ and, therefore, that the current $J^5_\mu(x)$ is conserved. This is a conclusion we now check by an explicit calculation of the expectation value of $\partial_\mu J^5_\mu(x)$ in the case of the action (2.2).

*Remarks.*

(i) For any regularization which is consistent with the hermiticity of $\gamma_5$

$$\left|\mathcal{Z}(A_\mu, \theta)\right|^2 = \det\,(\not{D}\not{D}^\dagger).$$

Therefore, an anomaly can appear only in the imaginary part of $\ln \mathcal{Z}$.

(ii) If the regularization is gauge invariant, $\mathcal{Z}(A_\mu, \theta)$ is also gauge invariant. Therefore, a possible anomaly will also be gauge invariant. One regularization scheme that has the required property is based on introducing regulator fields. But at least one regulator field must be an unpaired massive boson with fermion spin, to divide the fermion determinant by a factor $\det(\not{D} + \Lambda)$.

If this boson has a chiral charge, global chiral symmetry is broken by the mass $\Lambda$; if it has no chiral charge global chiral symmetry is preserved, and the determinant is independent of $\theta$ for $\theta(x)$ constant, but then the ratio of determinants is not invariant under local chiral transformations.

*General form of the anomaly.* The operator $\partial_\mu J^5_\mu(x)$ has dimension 4 and since a possible anomaly is a large momentum or short distance effect, $\langle \partial_\mu J^5_\mu(x) \rangle$ can only be a local function of $A_\mu$ of dimension 4. In addition parity implies that it is proportional to the completely antisymmetric tensor $\epsilon_{\mu\nu\rho\sigma}$. This determines $\left\langle \partial_\mu J^5_\mu(x) \right\rangle$ up to a multiplicative constant,

$$\left\langle \partial_\lambda J^5_\lambda(x) \right\rangle \propto e^2 \epsilon_{\mu\nu\rho\sigma} \partial_\mu A_\nu(x) \partial_\rho A_\sigma(x) \propto e^2 \epsilon_{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} \,,$$

$F_{\mu\nu}$ being the electromagnetic tensor. The possible anomaly is always gauge invariant.

To find the multiplicative factor, which is the only regularization dependent feature, it is sufficient to calculate the coefficient of term quadratic in $A$ in the expansion of $\left\langle\partial_\lambda J_\lambda^5(x)\right\rangle$ in powers of $A$. We define the three-point function:

$$\begin{aligned}\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2) &= \frac{\delta}{\delta A_\mu(p_1)}\frac{\delta}{\delta A_\nu(p_2)}\left\langle J_\lambda^5(k)\right\rangle\Big|_{A=0}, \qquad (2.7)\\ &= \frac{\delta}{\delta A_\mu(p_1)}\frac{\delta}{\delta A_\nu(p_2)} i\mathrm{tr}\left[\gamma_5\gamma_\lambda \not{D}^{-1}(k)\right]\Big|_{A=0}.\end{aligned}$$

$\Gamma^{(3)}$ is the sum of the two Feynman diagrams of Fig. 2.

The contribution of diagram (a) is

$$(\mathrm{a}) \mapsto \frac{e^2}{(2\pi)^4}\mathrm{tr}\left[\int \mathrm{d}^4q\,\gamma_5\gamma_\lambda(\not{q}+\not{k})^{-1}\gamma_\mu(\not{q}-\not{p}_2)^{-1}\gamma_\nu\not{q}^{-1}\right], \qquad (2.8)$$

and the contribution of diagram (b) is obtained by exchanging $p_1, \gamma_\nu \leftrightarrow p_2, \gamma_\nu$.

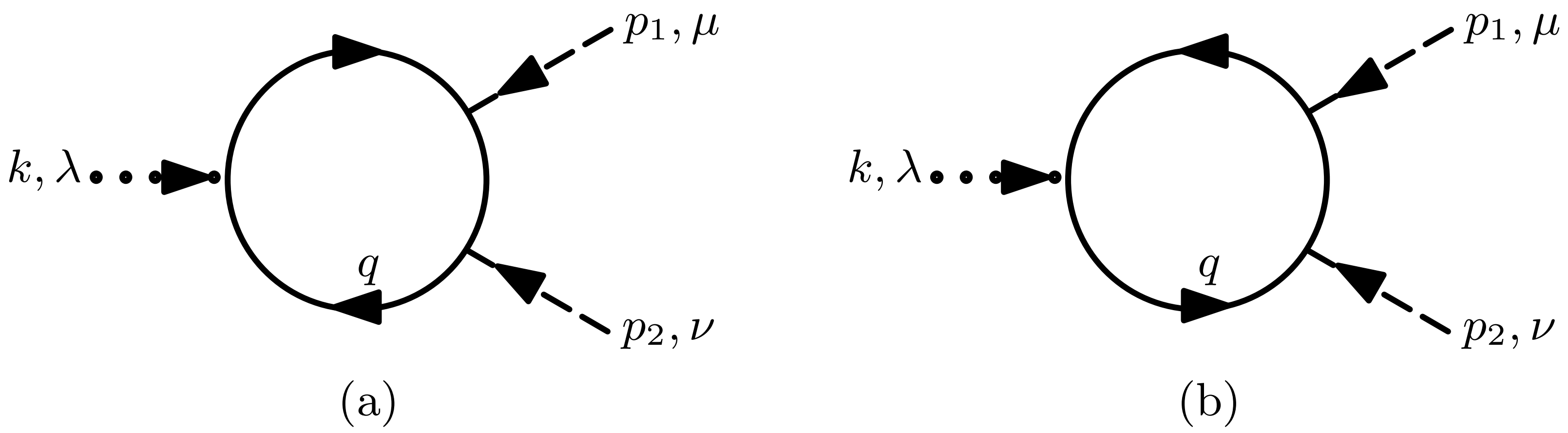


Fig. 2   Anomalous diagrams.

Power counting tells us that the function $\Gamma^{(3)}$ may have a linear divergence which, due to the presence of the $\gamma_5$ factor, must be proportional to $\epsilon_{\lambda\mu\nu\rho}$, symmetric in the exchange $p_1, \gamma_\nu \leftrightarrow p_2, \gamma_\nu$, and thus proportional to

$$\epsilon_{\lambda\mu\nu\rho}(p_1 - p_2)_\rho \,. \tag{2.9}$$

On the other hand, by commuting $\gamma_5$, one notices that $\Gamma^{(3)}$ is formally a symmetric function of the three sets of external arguments. A divergence breaks the symmetry between external arguments. Therefore, a symmetric regularization of the kind we adopt below leads to a finite result. The result is not ambiguous because a possible ambiguity again is proportional to (2.9).

In the same way, if the regularization is consistent with vector gauge invariance the WT identity

$$p_{1\mu}\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2)=0\,, \tag{2.10}$$

is satisfied. Applied to the divergent part it yields

$$-p_{1\mu}p_{2\rho}\epsilon_{\lambda\mu\nu\rho}=0\,,$$

a condition that is not satisfied. Therefore, the sum of the two diagrams is finite.

Different regularizations may still differ by finite quantities of the form (2.9) but again all regularizations consistent with vector gauge invariance must give the same answer.

Therefore, there are only two possibilities:

(i) The divergence $k_\lambda \Gamma^{(3)}_{\lambda\mu\nu}(k; p_1, p_2)$ in a regularization respecting the symmetry between the three arguments vanishes. Then $\Gamma^{(3)}$ is gauge invariant and the axial current is conserved.

(ii) The divergence of the symmetric regularization does not vanish. Then it is possible to add to $\Gamma^{(3)}$ a term proportional to (2.9) to restore gauge invariance but this term breaks the symmetry between external momenta: the axial current is not conserved, an anomaly is present.

*Divergence in the regularized theory*

The calculation can be done using one of the various gauge invariant regularizations, for example Pauli–Villars's regularization or dimensional regularization with $\gamma_5$ being defined as in dimension 4 and thus no longer anti-commuting with other $\gamma$ matrices. Instead we choose a regularization which preserves the symmetry between the three external arguments and global chiral symmetry, but breaks gauge invariance, replacing in the fermion propagator:

$$(\not{q})^{-1} \longmapsto (\not{q})^{-1} \rho(\varepsilon q^2),$$

where $\varepsilon$ is the regularization parameter ($\varepsilon \to 0$), $\rho(z)$ is a positive differentiable function such that $\rho(0) = 1$, and decreasing at least like $1/z$ for $z \to +\infty$.

Then the compatibility between current conservation and gauge invariance implies that $k_\lambda \Gamma^{(3)}_{\lambda\mu\nu}(k; p_1, p_2)$ vanishes.

It is convenient to consider directly the contribution $C^{(2)}(k)$ of order $A^2$ to $\langle k_\lambda J^5_\lambda(k)\rangle$ which sums the two diagrams:

$$C^{(2)}(k)=e^2\int \mathrm{d}^4p_1\,\mathrm{d}^4p_2\,A_\mu(p_1)A_\nu(p_2)\int\frac{\mathrm{d}^4q}{(2\pi)^4}\rho\big(\varepsilon(q+k)^2\big)\rho\big(\varepsilon(q-p_2)^2\big)\rho\big(\varepsilon q^2\big)$$
$$\times\,\mathrm{tr}\,\left[\gamma_5\not{k}(\not{q}+\not{k})^{-1}\gamma_\mu(\not{q}-\not{p}_2)^{-1}\gamma_\nu\not{q}^{-1}\right],$$

because the calculation then suggests how the method generalizes to arbitrary even dimensions. The calculation relies on the cyclic property of the trace and the *anticommutation* of $\gamma_5$.

We transform the expression, using the identity

$$(\not{q})^{-1}\not{k}(\not{q}+\not{k})^{-1}=(\not{q})^{-1}-(\not{q}+\not{k})^{-1},\tag{2.11}$$

and obtain

$$C^{(2)}(k)=e^2\int \mathrm{d}^4p_1\,\mathrm{d}^4p_2\,A_\mu(p_1)A_\nu(p_2t)\int\frac{\mathrm{d}^4q}{(2\pi)^4}\rho\big(\varepsilon(q+k)^2\big)\rho\big(\varepsilon(q-p_2)^2\big)$$
$$\times\,\rho\big(\varepsilon q^2\big)\mathrm{tr}\,\left\{\gamma_5\gamma_\mu(\not{q}-\not{p}_2)^{-1}\gamma_\nu\left[\not{q}^{-1}-(\not{q}+\not{k})^{-1}\right]\right\}.\tag{2.12}$$

We separate the two contributions in the right hand side. In the second contribution, proportional to $(\not{q}+\not{k})^{-1}$, we interchange $(p_1,\,\mu)$ and $(p_2,\,\nu)$ and shift $q \mapsto q+p_1$. Combining again the two contributions, one finds,

$$C^{(2)}(k) = \int \mathrm{d}^4 p_1\, \mathrm{d}^4 p_2\, A_\mu(p_1) A_\nu(p_2) \int \frac{\mathrm{d}^4 q}{(2\pi)^4} \rho\bigl(\varepsilon(q-p_2)^2\bigr)\rho\bigl(\varepsilon q^2\bigr)$$
$$\times \operatorname{tr}\left[\gamma_5 \gamma_\mu (\not{q}-\not{p}_2)^{-1} \gamma_\nu \not{q}^{-1}\right] \left[\rho\bigl(\varepsilon(q+k)^2\bigr) - \rho\bigl(\varepsilon(q+p_1)^2\bigr)\right].$$

We see that the two terms would cancel in the absence of regulators. This corresponds to the formal proof of current conservation. However, without regularization the integrals diverge and previous manipulations are not legitimate.

Here, instead, one finds a non-vanishing sum because the regulating factors which are different.

After evaluation of the trace, the sum becomes

$$C^{(2)}(k) = 4e^2 \int \mathrm{d}^4 p_1 \, \mathrm{d}^4 p_2 \, A_\mu(p_1) A_\nu(p_2) \int \frac{\mathrm{d}^4 q}{(2\pi)^4} \rho\big(\varepsilon(q-p_2)^2\big) \rho\big(\varepsilon q^2\big)$$
$$\times \, \epsilon_{\mu\nu\rho\sigma} \frac{p_{2\rho} q_\sigma}{q^2 (q-p_2)^2} \left[\rho\big(\varepsilon(q+p_1)^2\big) - \rho\big(\varepsilon(q+k)^2\big)\right].$$

Contributions coming from finite values of $q$ cancel in the $\varepsilon \to 0$ limit. Due to the cut-off, the relevant values of $q$ are of order $\varepsilon^{-1/2}$. One can, therefore, simplify the $q$ integrand:

$$\int \frac{\mathrm{d}^4 q}{(2\pi)^4 q^4} p_{2\rho} q_\sigma \rho^2(\varepsilon q^2) \rho'(\varepsilon q^2) \left[2\varepsilon q_\lambda (p_1 - k)_\lambda\right].$$

The identity:

$$\int \mathrm{d}^4 q \, q_\alpha q_\beta f(q^2) = \tfrac{1}{4} \delta_{\alpha\beta} \int \mathrm{d}^4 q \, q^2 f(q^2),$$

transforms the integral into

$$\tfrac{1}{2} p_{2\rho}(2p_1 + p_2)_\sigma \int \frac{\varepsilon \mathrm{d}^4 q}{(2\pi)^4 q^2} \rho^2(\varepsilon q^2) \rho'(\varepsilon q^2).$$

The remaining integral can be calculated explicitly (we recall $\rho(0) = 1$)

$$\int \frac{\varepsilon \mathrm{d}^4 q}{(2\pi)^4 q^2} \rho^2(\varepsilon q^2) \rho'(\varepsilon q^2) = \frac{1}{8\pi^2} \int_0^\infty \varepsilon q \mathrm{d}q\, \rho^2(\varepsilon q^2) \rho'(\varepsilon q^2) = -\frac{1}{48\pi^2},$$

and yields a result independent of the function $\rho$. One finally obtains

$$\left\langle k_\lambda J^5_\lambda(k) \right\rangle = -\frac{e^2}{12\pi^2} \epsilon_{\mu\nu\rho\sigma} \int \mathrm{d}^4 p_1\, \mathrm{d}^4 p_2\, p_{1\mu} A_\nu(p_1) p_{2\rho} A_\sigma(p_2). \tag{2.13}$$

From the definition (2.7), one concludes

$$k_\lambda \Gamma^{(3)}_{\lambda\mu\nu}(k; p_1, p_2) = \frac{e^2}{6\pi^2} \epsilon_{\mu\nu\rho\sigma} p_{1\rho} p_{2\sigma}\,.$$

This non-vanishing result implies that any definition of the determinant $\det \not{D}$ breaks at least either current conservation or gauge invariance.

Since gauge invariance is essential to the consistency of QED, one chooses to break current conservation. Exchanging arguments, one obtains the value of $p_{1\mu}\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2)$:

$$p_{1\mu}\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2) = \frac{e^2}{6\pi^2}\epsilon_{\lambda\nu\rho\sigma}k_\rho p_{2\sigma}\,.$$

By contrast, if we had used a gauge invariant regularization, the result for $\Gamma^{(3)}$ would have differed by a term $\delta\Gamma^{(3)}$ proportional to (2.9):

$$\delta\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2) = K\epsilon_{\lambda\mu\nu\rho}(p_1-p_2)_\rho\,.$$

The constant $K$ then is determined by the condition of gauge invariance

$$p_{1\mu}\left[\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2) + \delta\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2)\right] = 0\,,$$

which yields

$$p_{1\mu}\delta\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2) = -\frac{e^2}{6\pi^2}\epsilon_{\lambda\nu\rho\sigma}k_\rho p_{2\sigma} \;\Rightarrow\; K = e^2/(6\pi^2).$$

This gives an additional contribution to the divergence of the current

$$k_\lambda \delta\Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2) = \frac{e^2}{3\pi^2}\epsilon_{\mu\lambda\rho\sigma}p_{1\rho}p_{2\sigma}\,.$$

Therefore, in a QED-like gauge invariant field theory with massless fermions the axial current is not conserved: this is called the chiral anomaly. For any gauge invariant regularization one finds

$$k_\lambda \Gamma^{(3)}_{\lambda\mu\nu}(k;p_1,p_2) = \left(\frac{e^2}{2\pi^2} \equiv \frac{2\alpha}{\pi}\right)\epsilon_{\mu\nu\rho\sigma}p_{1\rho}p_{2\sigma}\,. \tag{2.14}$$

Equation (2.14) can be rewritten, after Fourier transformation, as a non-conservation equation for the axial current:

$$\partial_\lambda J^5_\lambda(x) = -i\frac{\alpha}{4\pi}\epsilon_{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}\,. \tag{2.15}$$

Since global chiral symmetry is not broken, the integral over the whole space of the anomalous term must vanish.

This condition is indeed verified since the anomaly can be written as a total derivative:

$$\epsilon_{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma} = 4\partial_\mu \left( \epsilon_{\mu\nu\rho\sigma} A_\nu \partial_\rho A_\sigma \right).$$

The space integral of the anomalous term depends only on the behaviour of the gauge field at boundaries, and this property indicates a relation between topology and anomalies.

Equation (2.15) also implies

$$\ln\det\left[\mathrm{e}^{i\gamma_5\theta(x)} \not{D} \mathrm{e}^{i\gamma_5\theta(x)}\right] = \ln\det\not{D} - i\frac{\alpha}{4\pi}\int \mathrm{d}^4x\,\theta(x)\epsilon_{\mu\nu\rho\sigma}F_{\mu\nu}(x)F_{\rho\sigma}(x) + O(\theta^2).$$

*Non-Abelian vector gauge theories and Abelian axial current*

We still consider an Abelian axial current but now in the framework of a non-Abelian gauge theory. The fermion fields transform non-trivially under a gauge group $G$ and $\mathbf{A}_\mu$ is the corresponding gauge field. The action is

$$\mathcal{S}(\bar\psi,\psi) = -\int \mathrm{d}^4 x\,\bar\psi(x)\,\not{\mathbf{D}}\,\psi(x)$$

with

$$\not{\mathbf{D}} = \not{\partial} + \not{\mathbf{A}}\,. \tag{2.16}$$

The axial current

$$J^5_\mu(x) = i\bar\psi(x)\gamma_5\gamma_\mu\psi(x),$$

is still gauge invariant. Therefore, no new calculation is needed; the result is completely determined by dimensional analysis, gauge invariance and the previous calculation which yields the term of order $\mathbf{A}^2$:

$$\partial_\lambda J^5_\lambda(x) = -\frac{i}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\,\mathbf{F}_{\mu\nu}\mathbf{F}_{\rho\sigma}\,, \tag{2.17}$$

in which $\mathbf{F}_{\mu\nu}$ is now the corresponding curvature. Again this expression must be a total derivative. One verifies:

$$\epsilon_{\mu\nu\rho\sigma} \operatorname{tr} \mathbf{F}_{\mu\nu} \mathbf{F}_{\rho\sigma} = 4\, \epsilon_{\mu\nu\rho\sigma} \partial_\mu \operatorname{tr} \left( \mathbf{A}_\nu \partial_\rho \mathbf{A}_\sigma + \tfrac{2}{3} \mathbf{A}_\nu \mathbf{A}_\rho \mathbf{A}_\sigma \right). \tag{2.18}$$

*Anomaly and eigenvalues of the Dirac operator*

We assume that the spectrum of $\not{\mathbf{D}}$, the Dirac operator in a non-Abelian gauge field (equation (2.16)), is discrete (enclosing temporarily the fermions in a box if necessary) and call $d_n$ and $\varphi_n(x)$ the corresponding eigenvalues and eigenvectors:

$$\not{\mathbf{D}} \varphi_n = d_n \varphi_n \,.$$

The eigenvalues are gauge invariant, because in a gauge transformation of unitary matrix $\mathbf{g}(x)$ the Dirac operator becomes

$$\not{\mathbf{D}} \mapsto \mathbf{g}^{-1}(x) \not{\mathbf{D}} \mathbf{g}(x) \;\Rightarrow\; \varphi_n(x) \mapsto \mathbf{g}(x) \varphi_n(x).$$

For a unitary or orthogonal group, the massless Dirac operator is anti-hermitian; therefore, the eigenvalues are imaginary and the eigenvectors orthogonal. In addition we choose them with unit norm.

The anticommutation $\not{D}\gamma_5 + \gamma_5\not{D} = 0$ implies

$$\not{D}\gamma_5\varphi_n = -d_n\gamma_5\varphi_n\,.$$

Therefore, either $d_n$ is different from zero, and $\gamma_5\varphi_n$ is an eigenvector of $\not{D}$ with eigenvalue $-d_n$, or $d_n$ vanishes. The eigenspace corresponding to the eigenvalue 0 then is invariant under $\gamma_5$ which can be diagonalized: the eigenvectors of $\not{D}$ can be chosen eigenvectors of definite chirality, that is, eigenvectors of $\gamma_5$ with eigenvalue $\pm 1$,

$$\not{D}\varphi_n = 0\,,\quad \gamma_5\varphi_n = \pm\varphi_n\,.$$

We denote by $n_+$ and $n_-$ the dimensions of the eigenspaces of positive and negative chirality, respectively.

We now consider the determinant of the operator $\not{\mathbf{D}} + m$ regularized by mode truncation (mode regularization):

$$\det{}_N(\not{\mathbf{D}} + m) = \prod_{n \le N} (d_n + m),$$

keeping the $N$ lowest eigenvalues of $\not{\mathbf{D}}$ (in modulus), with $N - n_+ - n_-$ even, in such a way that the corresponding subspace is $\gamma_5$ invariant.

The regularization is gauge invariant because the eigenvalues of $\not{\mathbf{D}}$ are gauge invariant.

Note that, in the truncated space, the trace of $\gamma_5$ is the index of the Dirac operator:

$$\operatorname{tr} \gamma_5 = n_+ - n_- \,. \tag{2.19}$$

It does not vanish if $n_+ \neq n_-$, a situation which endangers axial current conservation.

In a chiral transformation (2.4) with $\theta$ constant, the determinant of $(\not{D}+m)$ becomes

$$\det{}_N(\not{D}+m) \mapsto \det{}_N\left(\mathrm{e}^{i\theta\gamma_5}(\not{D}+m)\mathrm{e}^{i\theta\gamma_5}\right).$$

We now consider the various eigenspaces.

If $d_n \neq 0$ the matrix $\gamma_5$ is represented by the Pauli matrix $\sigma_1$ in the sum of eigenspaces corresponding to the two eigenvalues $\pm d_n$ and $\not{D}+m$ by $d_n\sigma_3+m$. The determinant in the subspace is then

$$\det\left(\mathrm{e}^{i\theta\sigma_1}(d_n\sigma_3+m)\mathrm{e}^{i\theta\sigma_1}\right) = \det \mathrm{e}^{2i\theta\sigma_1}\det(d_n\sigma_3+m) = m^2 - d_n^2,$$

because $\sigma_1$ is traceless.

In the eigenspace of vanishing eigenvalue $d_n = 0$ with positive chirality, of dimension $n_+$, $\gamma_5$ is diagonal with eigenvalue 1 and, thus,

$$m^{n_+} \mapsto m^{n_+}\mathrm{e}^{2i\theta n_+}.$$

Similarly, in the eigenspace $d_n = 0$ of chirality $-1$

$$m^{n_-} \mapsto m^{n_-} \mathrm{e}^{-2i\theta n_-} .$$

One concluses

$$\det_N \left(\mathrm{e}^{i\theta\gamma_5} (\not{\mathbf{D}} + m) \mathrm{e}^{i\theta\gamma_5}\right) = \mathrm{e}^{2i\theta(n_+ - n_-)} \det_N (\not{\mathbf{D}} + m).$$

The ratio of both determinants is independent of $N$. Taking the limit $N \to \infty$, one finds

$$\det \left[ \left(\mathrm{e}^{i\gamma_5\theta} (\not{\mathbf{D}} + m) \mathrm{e}^{i\gamma_5\theta}\right) (\not{\mathbf{D}} + m)^{-1} \right] = \mathrm{e}^{2i\theta(n_+ - n_-)} . \tag{2.20}$$

The left hand side of equation (2.20) is obviously 1 when $\theta = n\pi$, which implies that the coefficient of $2\theta$ in the right hand side must indeed be an integer.

The variation of $\ln\det(\not{D}+m)$,

$$\ln\det\left[\left(\mathrm{e}^{i\gamma_5\theta}(\not{D}+m)\mathrm{e}^{i\gamma_5\theta}\right)(\not{D}+m)^{-1}\right]=2i\theta(n_+-n_-),$$

at first order in $\theta$ is related to the variation of the action (2.2) and, thus, to the expectation value of the integral of the divergence of the axial current $\left\langle\int \mathrm{d}^4x\partial_\mu J^5_\mu(x)\right\rangle$. In the limit $m=0$, it is thus related to the space integral of the chiral anomaly (2.17):

$$-\frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\int \mathrm{d}^4x\,\mathrm{tr}\,\mathbf{F}_{\mu\nu}\mathbf{F}_{\rho\sigma}=n_+-n_-\,. \tag{2.21}$$

Concerning this result several comments can be made:

(i) The property that the integral (2.21) is quantized shows that the form of the anomaly is related to topological properties of the gauge field since the integral does not change when the gauge field is deformed continuously.

The integral of the anomaly over the whole space thus depends only on the behaviour at large distances of the curvature tensor $\mathbf{F}_{\mu\nu}$ and the anomaly must be a total derivative as equation (2.18) confirms.

(ii) Gauge field configurations exist for which the right hand side of equation (2.21) does not vanish, for example, instantons. We have shown above that if massless fermions are coupled to such gauge fields the determinant resulting from the fermion integration necessarily vanishes. This has some physical implications that are examined later.

(iii) One might be surprised that $\det \not{D}$ is not invariant under global chiral transformations. However, we have just established that when the integral of the anomaly does not vanish, $\det \not{D}$ vanishes. This explains that to give a meaning to the right hand side of equation (2.20) we have been forced to introduce a mass to find a non-trivial result. The determinant of $\not{D}$ in the subspace orthogonal to eigenvectors with vanishing eigenvalue, even in presence of a mass, is chiral invariant by parity doubling, but for $n_+ \neq n_-$ not the determinant in the eigenspace of eigenvalue zero because the trace of $\gamma_5$ does not vanish in the eigenspace (equation (2.19)). In the limit $m \to 0$ the complete determinant vanishes but not the ratio of determinants for different values of $\theta$ because the powers of $m$ cancel.

## Instantons in the $SU(2)$ gauge theory

Instantons are finite action solutions to Euclidean (imaginary time) field equations. They describe barrier penetration effects in the semi-classical limit.

We first consider pure gauge theories. Actually it is sufficient to consider the gauge group $SU(2)$ since a general theorem states that, for a Lie group containing $SU(2)$ as a subgroup, the instantons are those of the $SU(2)$ subgroup.

In $SO(3)$ notation the gauge field $\mathbf{A}_\mu$ is a vector and the gauge action reads

$$\mathcal{S}(\mathbf{A}_\mu) = \frac{1}{4g} \int \left[\mathbf{F}_{\mu\nu}(x)\right]^2 \mathrm{d}^4 x$$

with

$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + \mathbf{A}_\mu \times \mathbf{A}_\nu \,.$$

We define the dual of the tensor $\mathbf{F}_{\mu\nu}$ by

$$\tilde{\mathbf{F}}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}\mathbf{F}_{\rho\sigma}\,.$$

Then the inequality

$$\int \mathrm{d}^4x \left[\mathbf{F}_{\mu\nu}(x) \pm \tilde{\mathbf{F}}_{\mu\nu}(x)\right]^2 \geq 0\,,$$

implies

$$\mathcal{S}(\mathbf{A}_\mu) \geq |Q(\mathbf{A}_\mu)|/4g\,,$$

where $Q(\mathbf{A}_\mu)$ is an expression that appears in the expression of the axial anomaly (here written in $SO(3)$ notation):

$$Q(\mathbf{A}_\mu) = \int \mathrm{d}^4x\, \mathbf{F}_{\mu\nu} \cdot \tilde{\mathbf{F}}_{\mu\nu}\,. \tag{2.22}$$

We have verified that the quantity $\mathbf{F}_{\mu\nu} \cdot \tilde{\mathbf{F}}_{\mu\nu}$ is a pure divergence,

$$\mathbf{F}_{\mu\nu} \cdot \tilde{\mathbf{F}}_{\mu\nu} = \partial_\mu V_\mu$$

with

$$V_\mu = 2\epsilon_{\mu\nu\rho\sigma} \left[ \mathbf{A}_\nu \cdot \partial_\rho \mathbf{A}_\sigma + \tfrac{1}{3} \mathbf{A}_\nu \cdot (\mathbf{A}_\rho \times \mathbf{A}_\sigma) \right].$$

Therefore, the integral depends only on the behaviour of the gauge field at large distances and its values are quantized (equation (2.21)). The bound involves a topological charge, $Q(\mathbf{A}_\mu)$.

The finiteness of the action implies that the classical solution must asymptotically become a pure gauge, that is, with our conventions,

$$-\tfrac{1}{2} i \mathbf{A}_\mu \cdot \sigma = \mathbf{g}(x) \partial_\mu \mathbf{g}^{-1}(x) + O\left(|x|^{-2}\right) \quad |x| \to \infty\,,$$

in which $\sigma$ are Pauli matrices and $\mathbf{g}(x)$ is an element of $SU(2)$.

Since $SU(2)$ is topologically equivalent to $S_3$, one is now led to the study of the homotopy classes of mappings from $S_3$ to $S_3$, which are also classified by an integer, the winding number.

The simplest one to one mapping corresponds to an element $\mathbf{g}(x)$ of the form

$$\mathbf{g}(x) = \frac{x_4 + i\mathbf{x}\cdot\sigma}{r}, \quad r = (x_4^2 + \mathbf{x}^2)^{1/2}$$

and, thus,

$$A_m^i \underset{r\to\infty}{\sim} 2\,(x_4\delta_{im} + \epsilon_{imk}x_k)\,r^{-2}, \quad A_4^i = -2x_i r^{-2}.$$

It follows that

$$\int \mathrm{d}^4x\,\mathbf{F}_{\mu\nu}\cdot\tilde{\mathbf{F}}_{\mu\nu} = \int \mathrm{d}\Omega\,\hat{n}_\mu V_\mu = 32\pi^2,$$

in which $\mathrm{d}\Omega$ is the measure on the sphere and $\hat{n}_\mu$ the unit vector normal to the sphere.

Comparing this result with equation (2.21), one verifies that one has indeed found the minimal action solution. In general, one then expects

$$\int \mathrm{d}^4 x \, \mathbf{F}_{\mu\nu} \cdot \tilde{\mathbf{F}}_{\mu\nu} = 32\pi^2 n$$

and, therefore,

$$\mathcal{S}(\mathbf{A}_\mu) \geq 8\pi^2 |n|/g \, .$$

The equality, which corresponds to a local minimum of the action, is obtained for fields satisfying the self-duality equations

$$\mathbf{F}_{\mu\nu} = \pm \tilde{\mathbf{F}}_{\mu\nu} \, ,$$

which are first-order partial differential equations. The one-instanton solution, which depends on an arbitrary scale parameter $\lambda$, is

$$A^i_m = \frac{2}{r^2 + \lambda^2} \left( x_4 \delta_{im} + \epsilon_{imk} x_k \right) , \quad m = 1, 2, 3 \, , \quad A^i_4 = -\frac{2x_i}{r^2 + \lambda^2} . \quad (2.23)$$

*The semi-classical $\theta$ vacuum.* We now introduce the temporal gauge $\mathbf{A}_4 = 0$. The classical minima of the classical potential correspond to gauge field components $\mathbf{A}_i$, $i = 1, 2, 3$, which are pure gauge functions of the three space variables $x_i$:

$$-\tfrac{1}{2} i \mathbf{A}_m \cdot \sigma = \mathbf{g}(x_i) \partial_m \mathbf{g}^{-1}(x_i) \,.$$

The structure of the classical minima is related to the homotopy classes of mappings of the group elements $\mathbf{g}$ into compactified $\mathbb{R}^3$ (because $\mathbf{g}(x)$ goes to a constant for $|x| \to \infty$), that is, again of $S_3$ into $S_3$ and thus the semi-classical vacuum has a periodic structure like the cosine potential in quantum mechanics. One verifies that the gauge equivalent in the temporal gauge of the instanton solution (2.23) connects minima with different winding numbers. Therefore, to project onto a $\theta$-vacuum, one adds a term to the classical action of gauge theories:

$$\mathcal{S}_\theta(\mathbf{A}_\mu) = \mathcal{S}(\mathbf{A}_\mu) + \frac{i\theta}{32\pi^2} \int d^4 x \, \mathbf{F}_{\mu\nu} \cdot \tilde{\mathbf{F}}_{\mu\nu} \,. \qquad (2.24)$$

One can then integrate over all fields $\mathbf{A}_\mu$ without restriction. At least in the semi-classical approximation, the gauge theory depends on an additional parameter, the angle $\theta$. For non-vanishing values of $\theta$, the additional term violates CP conservation and is at the origin of the strong CP violation problem: if $\theta$ does not vanish, experimental bounds imply for $\theta$ unnaturally small values.

*Fermions in an instanton background.* Consider the QCD gauge action:

$$\mathcal{S}(\mathbf{A}_\mu, \bar{\mathbf{Q}}, \mathbf{Q}) = -\int \mathrm{d}^4x \left[\frac{1}{4e^2}\operatorname{tr}\mathbf{F}^2_{\mu\nu} + \sum_{f=1}^{N_f} \bar{\mathbf{Q}}_f \left(\not{\mathbf{D}} + m_f\right)\mathbf{Q}_f\right].$$

First, if the $\theta$ term in (2.24) contributes and one fermion field is massless, the Dirac operator has at least one vanishing eigenvalue and the determinant resulting from the fermion integration vanishes. Then instantons do not contribute to the field integral and the strong CP violation problem is solved.

However, such an hypothesis seems to be inconsistent with experimental data.

Moreover, if the instantons contribute, they solve the $U(1)$ problem, that is, the absence of a Goldstone boson associated with the almost spontaneous breaking of the axial $U(1)$ current.

## Physical application

*The solution of the $U(1)$ problem.* In a theory in which the quarks are massless and interact through a colour gauge group, the action has a chiral $U(N_{\mathrm{F}}) \times U(N_{\mathrm{F}})$ symmetry, in which $N_{\mathrm{F}}$ is the number of flavours. The spontaneous breaking of the chiral group to its diagonal subgroup $U(N_{\mathrm{F}})$ leads to expect $N_{\mathrm{F}}^2$ Goldstone bosons associated with the axial currents. From the preceding analysis we know that the axial current corresponding to the $U(1)$ Abelian subgroup has an anomaly. Of course the WT identities that imply the existence of Goldstone bosons correspond to constant group transformations and, therefore, involve only the space integral of the divergence of the current. Since the anomaly is a total derivative one might have expected the integral to vanish. However, non-Abelian gauge theories admit instanton solutions which give a periodic structure to the vacuum.

These instanton solutions correspond to gauge configurations that approach non-trivial pure gauges at infinity and give the set of discrete non-vanishing values one expects from equation (2.21) to the space integral of the anomaly (2.17). This indicates (but no satisfactory calculation of the instanton contribution has been performed) that for small, but non-vanishing, quark masses the $U(1)$ axial current is far from being conserved and, therefore, no light would-be Goldstone boson is generated. This observation resolves a long standing puzzle since experimentally no corresponding light pseudoscalar boson is found for $N_F = 2, 3$.

## Lattice gauge theories

Lattice gauge theories play a double role, they provide a regularization for perturbative gauge theories in the continuum and, moreover, the only known non-perturbative definition. They can be used for theoretical purpose and also can be studied by numerical methods.

We concentrate first on pure lattice gauge theories (without fermions). Physically, this means realistic properties of QCD cannot be determined, but one can still investigate one essential question:

Does the theory generate confinement, that is, a force between charged particles increasing at large distances, so that heavy quarks in the fundamental representation cannot be separated?

Other problems have also been discussed in this framework: for example, the appearance of massive group singlet bound states in the spectrum (gluonium), the question of a deconfinement transition at finite physical temperature in QCD by treating fermions in the quenched approximation.

Lattice gauge theories provide a lattice regularization of the continuum gauge theories: the low temperature or small coupling expansion of the lattice model is a regularized continuum perturbation theory.

However, other analytic results can be obtained in the high temperature or strong coupling limit and in the mean field approximation.

Discussing only pure gauge theories (without matter fields) on the lattice, one discovers that gauge theories have properties quite different from usual lattice models in statistical physics. Physical properties cannot be related to the behaviour of a local field and this forces to examine the behaviour of a non-local quantity, a functional of loops called hereafter Wilson's loop to distinguish between the confined and deconfined phases.

## Gauge invariance on the lattice

The construction of lattice gauge theories is based on an idea of parallel transport described in the first lecture.

We start from a model possessing a global (rigid) symmetry group $G$, and we want to make it gauge invariant.

To each site $i$ ($i$ represents the set of lattice coordinates) of a lattice, we associate a set of dynamic variables, $\{\psi_i, \bar{\psi}_i\}$, representing matter fields, on which acts a unitary representation $\mathcal{D}(G)$ of a Lie group $G$ (e.g., $SU(N)$):

$$\psi_{\mathbf{g}} = \mathbf{g}\psi\,, \qquad \mathbf{g} \in \mathcal{D}(G)\,.$$

A model is gauge invariant (local invariance) if it is invariant under independent group transformation on each lattice site $i$. For the $\psi, \bar{\psi}$-measure of integration as well as for all the terms in the lattice action which depend only on one site, global invariance implies local invariance.

Problems arise only with terms that connect different lattice sites.

Let us consider, for example, a term in the action of the form $\bar{\psi}_i\psi_j$, $i$ and $j$ being different sites on the lattice. Such a term is invariant under global but not local transformations:

$$\bar{\psi}_i\psi_j \mapsto \bar{\psi}_i\mathbf{g}_i^\dagger\,\mathbf{g}_j\psi_j .$$

To render it invariant it is necessary to introduce a new dynamic variable, a matrix $\mathbf{U}_{ij}$ belonging to the representation $\mathcal{D}(G)$, which depends on the two sites $i, j$ and transforms like

$$\mathbf{U}_{ij} \mapsto \mathbf{g}_i\mathbf{U}_{ij}\mathbf{g}_j^{-1}\,. \tag{3.1}$$

Then, the quantity

$$\bar{\psi}_i\mathbf{U}_{ij}\psi_j \tag{3.2}$$

is gauge invariant. Moreover, if $\mathbf{U}_{ij}$ and $\mathbf{U}_{jk}$ are two matrices transforming with the rule (3.2), then the product of matrices $\mathbf{U}_{ij}\mathbf{U}_{jk}$ transforms like

$$\mathbf{U}_{ij}\mathbf{U}_{jk} \mapsto \mathbf{g}_i\mathbf{U}_{ij}\mathbf{U}_{jk}\mathbf{g}_k^{-1}\,. \tag{3.3}$$

In the transformation (3.1), one recognizes the transformation of a parallel transporter. In the continuum, a parallel transporter depends not only on the end-points $i, j$ but also on the curve joining them. Moreover, in a local field theory one needs only transport along infinitesimal curves which can be expressed in terms of a gauge field or connection, element of the representation of the Lie algebra.

On the lattice curves follow links, the segments which connect adjacent sites. The minimum displacement is a link and two arbitrary lattice sites can be joined by a path formed of links of the lattice. As a consequence of the composition rule (3.3), one can thus take as dynamic variables elements $\mathbf{U}_\ell$ of the group representation associated with parallel transport along oriented links of the lattice, which transform like

$$\mathbf{U}_\ell \equiv \mathbf{U}_{ab} \mapsto \mathbf{g}_a \mathbf{U}_\ell \mathbf{g}_b^{-1} ,$$

where the link $\ell$ goes from site $b$ to adjacent site $a$.

It is consistent with the transformation law to choose

$$\mathbf{U}_{ba} = \mathbf{U}_{ab}^{-1} . \tag{3.4}$$

Then, one can express a general parallel transporter $\mathbf{U}[C(i,j)]$ depending on a curve $C$ on the lattice as a product of link variables along the path $C$ joining $j$ to $i$:

$$\mathbf{U}[C(i,j)] = \prod_{\text{links } \ell \in \mathrm{C(i,j)}} \mathbf{U}_\ell \, ,$$

where the product is ordered along the path.

*Relation with the continuum formulation: the QED Abelian example.* In continuum field theory, in the Abelian $U(1)$ example, we have already explicitly constructed the parallel transporter which is an element of the $U(1)$ group. In terms of the gauge field $A_\mu$, it reads

$$U[C(x,y)] = \exp\left[-ie \oint_C A_\mu(s) \mathrm{d}s_\mu\right] .$$

Indeed in a gauge transformation a charged field $\psi$ and the gauge field transform like

$$\psi(x) \mapsto \mathrm{e}^{i\Lambda(x)}\psi(x), \quad A_\mu(x) \mapsto A_\mu(x) - \frac{1}{e}\partial_\mu \Lambda(x),$$

and thus

$$e \oint_C A_\mu(s)\mathrm{d}s_\mu \mapsto e \oint_C A_\mu(s)\mathrm{d}s_\mu - \Lambda(y) + \Lambda(x).$$

The transformation of $U[C(x,y)]$ is then

$$U[C(x,y)] \mapsto \mathrm{e}^{i\Lambda(y)-i\Lambda(x)} U[C(x,y)].$$

*The non-Abelian case.* In the non-Abelian case, the explicit relation is more complicated because the gauge field $\mathbf{A}^\alpha_\mu(x)t_\alpha$ is an element of the Lie algebra of $G$ and the matrices representing the field at different points do not commute. It can be formally written as (P means path-ordered integral)

$$\mathbf{U}(x,y) = \mathrm{P}\left\{\exp\left[\oint_C \mathbf{A}^\alpha_\mu(s)t^\alpha \mathrm{d}s_\mu\right]\right\}.$$

## The pure gauge theory

We now discuss the pure gauge theory and its formal continuum limit as obtained from a low temperature, strong coupling expansion.

*Gauge invariant action and partition function*

First, one must construct a gauge invariant interaction for the link variables $\mathbf{U}_{ij}$. It follows from the transformation (3.1) that only the traces of the products of $\mathbf{U}$'s along closed loops are gauge invariant. On a hypercubic lattice, the shortest loop is a square, called hereafter a plaquette. In what follows, we thus consider a pure gauge action of the form ($i, j, k, l$ form a square on the lattice)

$$\mathcal{S}(\mathbf{U}) = -\beta_p \sum_{\text{all plaquettes}} \operatorname{tr} \mathbf{U}_{ij}\mathbf{U}_{jk}\mathbf{U}_{kl}\mathbf{U}_{li}\,, \tag{3.5}$$

in which $\beta_p$ is the plaquette coupling (which is proportional to $1/g^2$).

The appearance of products of parallel transporters along closed loops is not surprising since the pure gauge action of the continuum theory is associated with infinitesimal transport along a closed loop.

Note that each plaquette appears with both orientations in such a way that the sum is real when the group is unitary.

*The quantum partition function.* We can then write the partition function corresponding to the action (3.5) as

$$\mathcal{Z} = \int \prod_{\text{links}\{ij\}} \mathrm{d}\mathbf{U}_{ij}\, \mathrm{e}^{-\mathcal{S}(\mathbf{U})}. \tag{3.6}$$

The integration measure is the group invariant (de Haar) measure associated with the group $G$. In contrast with continuum gauge theories, the expression (3.6) is well-defined on the lattice (at least as long as the volume is finite) because the group is compact and thus the volume of the group is finite.

Therefore, gauge fixing is not required and a *completely gauge invariant formulation* of the theory is possible.

*Low coupling analysis*

We first want to understand the precise connection between the lattice theory (3.6) and the continuum field theory. For this purpose, we investigate the lattice theory at low coupling, that is, at large positive $\beta_p$. In this limit, the partition function is dominated by minimal action configurations.

Let us show that the minimum of the action corresponds to matrices $\mathbf{U}$ gauge transform of the identity. We start from a first plaquette 1234. Without loss of generality, we can set

$$\mathbf{U}_{12} = \mathbf{g}_1^{-1}\mathbf{g}_2\,, \qquad \mathbf{g}_1,\, \mathbf{g}_2 \in \mathcal{D}(G).$$

The matrix $\mathbf{g}_1$ is arbitrary and $\mathbf{g}_2$ is calculated from $\mathbf{U}_{12}$ and $\mathbf{g}_1$. Then, we can also set

$$\mathbf{U}_{23} = \mathbf{g}_2^{-1}\mathbf{g}_3\,, \qquad \mathbf{U}_{34} = \mathbf{g}_3^{-1}\mathbf{g}_4\,.$$

These relations define first $\mathbf{g}_3$, then $\mathbf{g}_4$. The minimum of the action is obtained when the real part of all traces is maximum, that is, when the products of the group elements on a plaquette are 1. In particular,

$$\mathbf{U}_{12}\mathbf{U}_{23}\mathbf{U}_{34}\mathbf{U}_{41} = \mathbf{1}\,,$$

which yields

$$\mathbf{U}_{41} = \mathbf{g}_4^{-1}\mathbf{g}_1\,.$$

If we now take an adjacent plaquette the argument can be repeated for all links but one, which has already been fixed. In this way, one can show that the minimum of the action is a pure gauge. Thus, when the coupling constant $\beta_p$ becomes very large, all group elements are constrained to stay, up to a gauge transformation, close to the identity (in a finite volume with consistent boundary conditions).

From this analysis, one learns that the minimum of the lattice action is highly degenerate at low coupling, since it is parametrized by a gauge transformation, which corresponds to a finite number of degrees of freedom per site. This unusual property of lattice gauge theories corresponds to the property that the gauge action in classical mechanics determines the motion only up to a gauge transformation. To perform a low coupling expansion, it becomes necessary to 'fix' the gauge in order to sum over all minima.

*Low coupling expansion.* We choose a gauge such that the minimum of the action corresponds to all matrices $\mathbf{U} = \mathbf{1}$. At low coupling, the matrices $\mathbf{U}$ are then close to the identity:

$$\mathbf{U}(x, x + an_\mu) = 1 - a\mathbf{A}_\mu(x) + O\left(a^2\right),$$

in which $a$ is the lattice spacing, $x$ the point on the lattice, and $n_\mu$ the unit vector in the direction $\mu$.

In the matrix $\mathbf{A}_\mu(x)$ we recognize the connection or gauge field. One can then expand the lattice action for small fields.

At leading order, one finds

$$\sum_{\mu,\nu,x} \operatorname{tr} \mathrm{e}^{-a^2 \mathbf{F}_{\mu\nu}(x)} - \operatorname{tr} \mathbf{1} = a^4 \sum_{\mu,\nu,x} \operatorname{tr} \mathbf{F}^2_{\mu\nu}(x) + O\left(a^6\right),$$

where $\mathbf{F}_{\mu\nu}(x)$ is the curvature tensor

$$\mathbf{F}_{\mu\nu}(x) = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu + [\mathbf{A}_\mu, \mathbf{A}_\nu] + O\left(a\right).$$

This result shows that the leading term of the small field expansion of the plaquette action (3.5) is the standard gauge action. The relation between $\beta_p$ and the bare coupling constant $e_0$ of continuum gauge theories is thus

$$a^4 \beta_p \sim g_0^{-2}\,. \tag{3.7}$$

Therefore, the low coupling expansion, in a fixed gauge, of lattice gauge theories indeed provides a regularization of continuum gauge theories.

We have here discussed only the pure gauge action, but the generalization to matter fields is simple.

Higher order terms in the small field expansion yield additional interactions needed to maintain gauge invariance on the lattice. This is not surprising: we have already shown that the gauge invariant extension of Pauli–Villars's regularization also introduces additional interactions.

## Wilson's loop and confinement

The form of the RG $\beta$-function shows that gauge theories are asymptotically free in four dimensions, which means that the origin in the coupling constant space is an UV fixed point and also implies that the effective interaction increases at large distance. Therefore, the spectrum of a non-Abelian gauge theory cannot be determined from perturbation theory. To explain the non-observation of free quarks, it has been conjectured that the spectrum of the symmetric phase consists only in neutral states, that is, states which are singlets for the group transformations.

To discuss the confinement problem, it has been suggested by Wilson to study, in pure gauge theories, a gauge invariant non-local quantity, the energy of the vacuum in presence of largely separated static charges. We thus first examine this quantity in pure Abelian gauge theories, in which, in the continuum, all calculations can be done explicitly.

*Wilson's loop in continuum Abelian gauge theories*

In continuum field theory, in order to calculate the average energy, it is necessary to introduce the gauge Hamiltonian, and, therefore, convenient to work in the temporal gauge. We can construct a wave function for two static point-like charges, in the temporal gauge:

$$\psi(A) = \exp\left[-ie \oint_{C_0} A_i(s) \mathrm{d}s_i\right],$$

in which the charges are located at both ends of the curve $C_0$.

By evaluating the behaviour for large time $T$ of the matrix element

$$W(C_0) = \left\langle \psi \left| \mathrm{e}^{-HT} \right| \psi \right\rangle,$$

in which $H$ is the gauge Hamiltonian in the temporal gauge, one obtains the energy $E(C_0)$ of the vacuum in presence of static charges:

$$W(C_0) \underset{T\to\infty}{\sim} \mathrm{e}^{-TE(C_0)}.$$

If the charges are separated by a distance $R$, one expects $E$ to depend only on $R$ and not on $C_0$.

The loop functional $W(C_0)$ can be calculated from a field integral:

$$W(C_0) = \left\langle \exp\left[-ie \oint_{C_0'} A_\mu(s)\mathrm{d}s_\mu\right]\right\rangle .$$

$C_0'$, which is now defined in space and time, is the union of two curves, which coincide with $C_0$ at time 0, and with $-C_0$ at time $T$, respectively. The expectation value here means average over gauge field configurations.

Since in the temporal gauge the time component of $A_\mu$ vanishes, we can add to $C_0'$ two straight lines in the time direction which join the ends of the curves $C_0(t = 0)$ and $C_0(t = T)$. $W(C_0)$ then becomes a functional of a closed loop $C$ (see figure 3):

$$W(C_0) \equiv W(C) = \left\langle \exp\left[-ie \oint_{C} A_\mu(s)\mathrm{d}s_\mu\right]\right\rangle . \tag{3.8}$$

The advantage of the representation (3.8) is that it is explicitly gauge invariant since it is the expectation value of the parallel transporter corresponding to a closed loop in space–time.

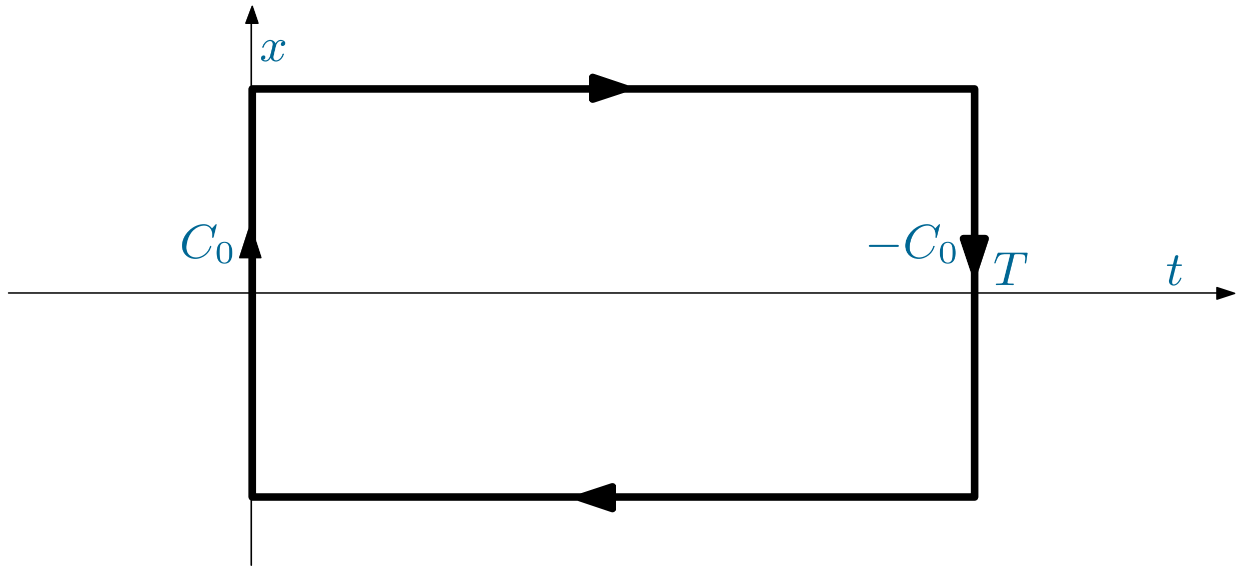


Fig. 3 The loop $C$.

The question of confinement is related to the behaviour of the energy $E$ when the separation $R$ between charges becomes large.

In a pure Abelian gauge theory, in the continuum, which is a free field theory, the expression (3.8) can be evaluated explicitly. To simplify calculations we take for $C_0$ also a straight line and use Feynman's gauge. The quantity $W(C)$ then is given by

$$W(C) = \int [\mathrm{d}A_\mu] \exp\left[-\mathcal{S}(A_\mu) + \int \mathrm{d}^d x\, J_\mu(x) A_\mu(x)\right]$$

with

$$\mathcal{S}(A_\mu) = \frac{1}{2} \int \mathrm{d}^d x\, [\partial_\mu A_\nu(x)]^2$$

and

$$J_\mu(x) = -ie \oint_C \delta(x - s) \mathrm{d}s_\mu \,.$$

The result is

$$\ln W(C) = -\frac{\Gamma(d/2-1)}{8\pi^{d/2}} \mathrm{e}^2 \oint_{C\times C} \mathrm{d}\mathbf{s}_1 \cdot \mathrm{d}\mathbf{s}_2 \,|\mathbf{s}_1 - \mathbf{s}_2|^{2-d} \,. \qquad (3.9)$$

The integral in the right hand side exhibits a short distance singularity, which requires a short distance cut-off. Moreover, to normalize the right hand side of equation (3.9), we divide it by $W(C)$ taken its value at a fixed distance $R = a$. We now write the integrals more explicitly:

$$\oint_{C\times C} \frac{\mathrm{d}\mathbf{s}_1 \cdot \mathrm{d}\mathbf{s}_2}{2\left|\mathbf{s}_1 - \mathbf{s}_2\right|^{d-2}} = \int_0^T |u-t|^{2-d}\, \mathrm{d}u\, \mathrm{d}t + \int_0^R |x-y|^{2-y}\, \mathrm{d}x\, \mathrm{d}y$$
$$-\int_0^R \left[(x-y)^2 + T^2\right]^{1-d/2} \mathrm{d}x\, \mathrm{d}y - \int_0^T \left[(t-u)^2 + R^2\right]^{1-d/2} \mathrm{d}t\, \mathrm{d}u\,.$$

The first term in the right hand side is cancelled by the normalization. The second term is independent of $T$ and, therefore, negligible for large $T$. The third term decreases with $T$ for $d > 2$, which we now assume.

Only the last term increases with $T$:

$$\int_0^T \left\{\left[(t-u)^2+R^2\right]^{1-d/2}-\left[(t-u)^2+a^2\right]^{1-d/2}\right\} \mathrm{d}t\,\mathrm{d}u$$
$$\sim \sqrt{\pi}\,\frac{\Gamma\big((d-3)/2\big)}{\Gamma(d/2-1)}\left(R^{3-d}-a^{3-d}\right)T\,.$$

Therefore, the vacuum energy $E(R)$ in presence of the static charges has the form

$$E(R)-E(a)=\frac{e^2}{4\pi^{(d-1)/2}}\Gamma\big((d-3)/2\big)\left(a^{3-d}-R^{3-d}\right).$$

One recognizes the Coulomb potential between two charges.

For $d \le 3$, the energy of the vacuum increases without bound when the charges are separated, and free charges cannot exist.

For $d=3$, the potential increases logarithmically.

For $d=2$, the Coulomb potential increases linearly with distance.

In more general situations, the method that we have used above to determine the energy is complicated because we have to take the large $T$ limit first and then evaluate the large $R$ behaviour. It is more convenient to take a square loop, $T = R$, and evaluate the large $R$ behaviour of $W(C)$. Here, one obtains

$$\begin{aligned}
&\ln W\big[C(R)\big] - \ln W\big[C(a)\big] \\
&\quad = \frac{1}{2\pi^{d/2}}\Gamma(d/2-1)e^2 \left\{ \int_0^R \left[(u-t)^2 + R^2\right]^{1-d/2} \mathrm{d}u\,\mathrm{d}t \right. \\
&\quad \left. - \int_0^a \left[(u-t)^2 + a^2\right]^{1-d/2} \mathrm{d}u\,\mathrm{d}t - \int_a^R |u-t|^{2-d} \mathrm{d}u\,\mathrm{d}t \right\}.
\end{aligned}$$

For $d > 3$, dimensions in which the Coulomb potential decreases, the right hand side is dominated by terms that correspond to the region $|s_1 - s_2| \ll R$ in equation (3.9):

$$\ln W\left[C(R)\right] - \ln W\left[C(a)\right] \sim \text{const.} \times R\,.$$

This is called the perimeter law since $\ln W(C)$ is proportional to the perimeter of $C$ and is, therefore, relevant to the $d = 4$ Coulomb phase.

Instead for $d \leq 3$, $\ln W(C)$ increases as $R^{4-d}$. The reason is that two charges separated on $C$ by a distance of order $R$, feel a potential of order $R^{d-3}$.

In particular for $d = 2$, $\ln W(C)$ increases like $R^2$, that is, like the area of the surface enclosed by $C$: this is the area law expected in confinement situations.

*Non-Abelian gauge theories*

In the temporal gauge the wave function corresponding to two opposite point-like static charges is also related to a parallel transporter along a curve joining the charges.

The same arguments as in the Abelian case, show that the expectation value of the operator $\mathrm{e}^{-TH}$ in the corresponding state is given by the average, in the sense of the functional integral, of the parallel transporter along a closed loop:

$$W(C) = \left\langle \mathrm{P} \exp \left[ -i \oint_C \mathbf{A}_\mu(s) \mathrm{d}s_\mu \right] \right\rangle ,$$

in which we recall that the symbol P means path ordering since the matrices $\mathbf{A}_\mu(s)$ at different points do not commute.

If we calculate $W(C)$ in perturbation theory, we find of course at leading order the same results as in the Abelian case. However, we know from renormalization group, that we cannot trust perturbation theory at large distances. Therefore, to get a qualitative idea about the phase structure, one can use the lattice model to calculate $W(C)$ in the large coupling or high temperature limit $\beta_p \to 0$.

*Strong coupling expansion for Wilson's loop.* We here assume that the group we consider has a non-trivial centre. We shall take the explicit example of gauge elements on the lattice belonging to the fundamental representation of $SU(N)$ (whose centre is $\mathbb{Z}_N$, with elements the identity multiplied by roots $z$ of unity, $z^N = 1$).

We calculate $W(C)$ by expanding the integrand in expression (3.6) in powers of $\beta_p$. We choose for simplicity for the loop $C$ a rectangle although the generalization to other contours is simple.

Any non-vanishing contribution must be invariant by the change of variables $U_\ell \mapsto z_\ell U_\ell$, where $z_\ell$ belongs to the centre. Let us consider one link belonging to the loop and multiply the corresponding link variable $\mathbf{U}(x, x+an_\mu)$ by $z_0$. We now multiply all link variables $\mathbf{U}(x+y, x+y+an_\mu)$, which are obtained by a translation $y$ in the hyperplane perpendicular to $n_\mu$, by $z_y$. Another link belonging to the loop belongs to the set but with opposite orientation.

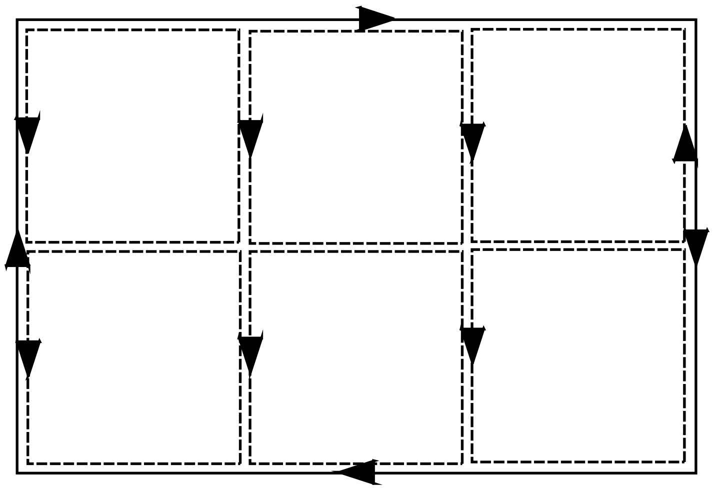

Fig. 4 The inside of loop $C$ covered with plaquettes: Area law at strong coupling

Plaquettes involving such variables involve them in pairs. For a result to be invariant and thus non-vanishing, the number of times each link variable appears in the direction $n_\mu$ minus the number of times it appears in the direction $-n_\mu$ must vanish (mod $N$). Thus, we start adding plaquettes to satisfy this condition at point $x$. However, the addition of one plaquette does not change the total difference between the numbers of links in the $+n_\mu$ and $-n_\mu$ directions.

Therefore, always at least one condition remains unsatisfied until the plaquettes reach the other link of the loop. We can then repeat the arguments for the remaining links of the loop and the new non integrated remaining links of the plaquettes. The number of required plaquette variables to get a non-vanishing result, is at least equal to the area of the rectangle, the minimal area surface having the loop as boundary (see figure 4). We can then perform the integrations which are just factorized group integrations. In this way, we get a contribution to $W(C)$ proportional to $(\beta_p)^A$, in which $A$ is the number of plaquettes. The largest contribution corresponds to plaquettes covering the minimal area surfaces bounded by the loop. It is indeed obtained by covering the rectangle with plaquettes in such a way that each link variable appears only twice in either orientation. For a rectangular loop $R \times T$, one gets

$$W(C) \sim \mathrm{e}^{RT \ln \beta_p}. \tag{3.10}$$

This result indicates that the potential between the static charges is linearly rising at large distance. Static charges creating the loop cannot be screened by the gauge field, otherwise one would have found a perimeter law.

*Remarks.*

(i) If the centre is trivial, it is possible to form a tube along the loop and this implies a perimeter law. If, for example, the group is $SO(3)$, in the decomposition of a product of two spin 1 representations, we again find a spin 1 which can be coupled to a third spin 1 to form a scalar. Thus, two plaquettes can be glued to the same link of the loop without constraint on the orientation of the plaquette.

(ii) The asymptotic form (3.10) is also valid for the Abelian $U(1)$ lattice gauge theory. Therefore, in four dimensions, Wilson's loop has a perimeter law in the weak coupling expansion and an area law at large coupling. One thus expects a phase transition between a low coupling Coulomb phase, described by a free field theory, and a strong coupling confined phase.

*The string tension.* The coefficient in front of the area

$$\sigma(\beta_p) \underset{\beta_p \to 0}{\sim} -\ln \beta_p \,,$$

is called the string tension. If no phase transition occurs when $\beta_p$ varies from zero to infinity, the gauge theory leads to confinement. In this case, the behaviour of the string tension for $\beta_p$ small is predicted by the renormalization group. Since $\sigma$ has the dimension of a mass squared, one finds

$$\sigma(g_0) \sim (g_0^2)^{-\beta_2/\beta_3^2} \exp\left(-1\,/\beta_2 g_0^2\right) . \tag{3.11}$$

in which $g_0^2$ is related to $\beta_p$ by equation (3.7) and $\beta_2$, $\beta_3$ are two first coefficients of the RG $\beta$-function. A physical quantity relevant to the continuum limit can then be obtained by dividing $\sqrt{\sigma}$ by its asymptotic behaviour. Let us define $\Lambda_L$ as

$$\Lambda_L = a^{-1}(\beta_2 g_0^2)^{-\beta_3/2\beta_2^2} \exp\left(-1/2\beta_2 g_0^2\right) .$$

Then $\Lambda_L / \sqrt{\sigma}$ has a continuum limit. When one calculates $\sigma$ by non-perturbative lattice methods, the verification of the scaling behaviour (3.11) indicates that the result is relevant to the continuum field theory and not only a lattice artifact.

It is possible to systematically expand $\sigma$ in powers of $\beta_p$. The possibility of verifying that confinement is realized in the continuum limit, depends on the possibility of analytically continuing the strong coupling expansion up to the origin. Unfortunately, theoretical arguments lead to believe that, independently of the group, the string tension is affected by a singularity associated with the roughening transition, transition which, however, is not related to bulk properties. At strong coupling, the contributions to the string tension come from smooth surfaces. When $g_0^2$ decreases ($\beta_p$ increases), one passes through a critical point $g_{0R}^2$, after which the relevant surfaces become rough. At the singular coupling $g_{0R}^2$, the string tension does not vanish but has a weak singularity.

Still at this point the strong coupling expansion diverges. Therefore, it is impossible to extrapolate to arbitrarily small coupling. The usefulness of the strong coupling expansion then depends on the position of the roughening transition with respect of the onset of weak coupling behaviour. Notice that numerically in the neighbourhood of the roughening transition, rotational symmetry is approximately restored (at least at large enough distance).

One can also calculate other quantities which are associated to bulk properties, and are, therefore, not affected by roughening singularities, such as the free energy (the connected vacuum amplitude) or the plaquette–plaquette correlation function. However, even for these quantities the extrapolation is not easy because the transition between strong and weak coupling behaviours is in general very sharp. From the numerical point of view, it seems that the plaquette–plaquette correlation function is the most promising case for strong coupling expansion.

*Remark.* The potential between static charges increases linearly in the same way as the Coulomb potential in one space dimension. This leads to the following physical picture: in QED the gauge field responsible of the potential has no charge and propagates essentially like a free field isotropically in all space directions. Conservation of flux on a sphere then yields the $R^{2-d}$ force between the charges. However, in the non-Abelian case the attractive force between the gauge particles generates instead a flux tube between static charges in such a way that the force remains the same as in one space dimension.

*Numerical methods: Computer simulations.* We will not describe the numerical methods which have been used in lattice gauge theories. In pure gauge theories, the existence of phase transitions has been investigated for many lattice actions. For the gauge group $SU(3)$, relevant to the physics of Strong Interactions, the string tension has been carefully measured, the plaquette–plaquette correlation function has been studied to determine the mass of low lying gluonium states. Finally, calculations have been performed at finite physical temperature, that is, on a 3+1 dimensional lattice in the limit in which the size of the lattice remains finite in one dimension, this size being related to the temperature. In this way, the temperature of a deconfinement transition has been determined.

*Fermions in numerical simulations.* One important qualitative feature of Strong Interaction physics is the approximate spontaneous breaking of chiral symmetry. However, non-trivial problems arise when one tries to construct a chiral invariant lattice action. One has the choice only between writing an action which is not explicitly chiral symmetric and in which one tries to restore chiral symmetry by adjusting the fermion mass term (Wilson's fermions), writing a chiral symmetric action with too many fermions (staggered or Kogut–Susskind fermions), or, as it has been more recently discovered various Dirac operators satisfying the Ginsparg–Wilson relation, called overlap or domain wall fermions. In the latter solution, several implementations can be interpreted as adding for the fermions an extra space dimension, which increases the already difficult computer problem.

Indeed, an important practical difficulty also arises with fermions: because it is impossible to simulate numerically fermions, it is necessary to integrate over fermions explicitly. This generates an effective gauge field action which contains a contribution proportional to the fermion determinant and is, therefore, no longer local. The speed of numerical methods crucially depends on the locality of the action. This explains that most numerical simulations with fermions have been up to now performed in the so-called quenched approximation in which the determinant is neglected. This approximation corresponds to the neglect of all fermion loops and bears some similarity with the eikonal approximation. In this approximation, the approximate spontaneous breaking of chiral symmetry has been verified by measuring the decrease of the pion mass for decreasing quark masses. Owing to the difficulty of the problem, the numerical study of the effect of dynamical fermions for realistic lattice sizes and close enough to the chiral limit has begon only recently.